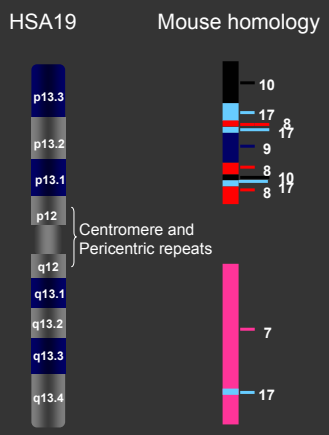
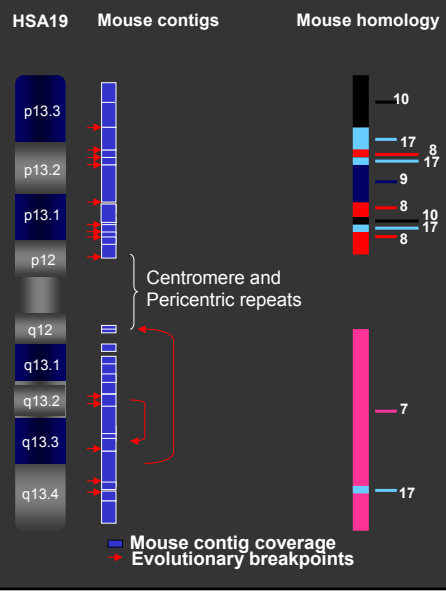
Human Chromosome 19:

- ~65 Mb length
  - ~ 1300 genes
  - ~17 Mb centromere + pericentromeric, ~2 Mb additional gene “deserts”
  - 46 Mb contains most genes (~1gene /36 kb)
- 57 Mb contiguous clone map with 7 gaps
  - ~43 Mb finished sequence, 14 Mb draft
  - targeted for finishing ~7/02
- 15 homology segments related to mouse chromosomes 7, 8, 9, 10 and 17





Initial analysis focused on three major questions:

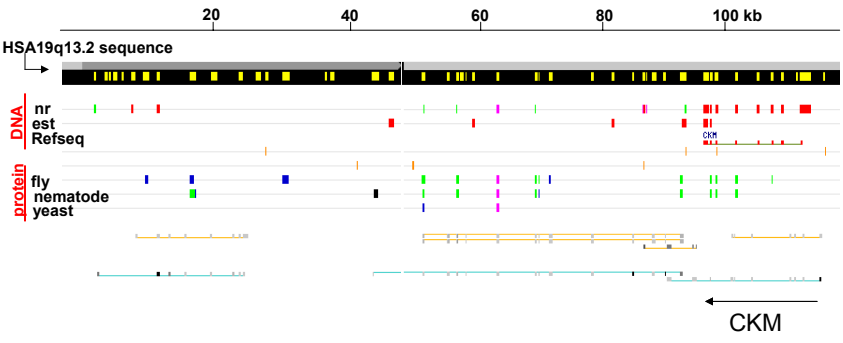
- Human sequence annotation:
  - **what is the value of comparative alignments for gene finding and functional-element identification?**
- Chromosome evolution:
  - **have chromosome rearrangements that distinguish human and mouse chromosomes occurred at random or specific sites?**
- Gene evolution:
  - **How do primate and rodent gene sets compare?**

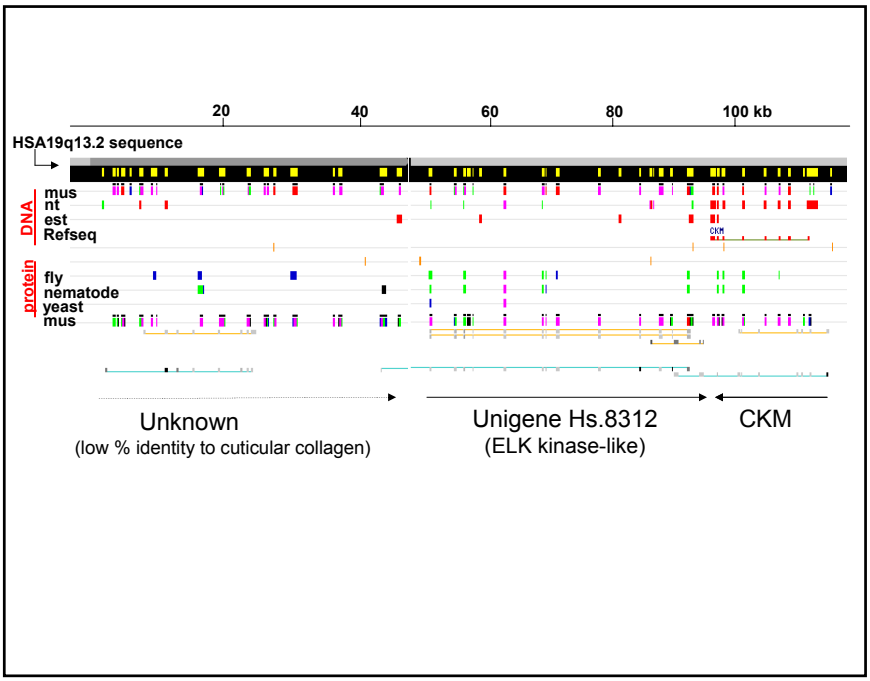
Value of mouse comparison as a sequence-annotation strategy

- What did we gain by sequencing mouse?
  - Identification of many new candidate exons for partially sequenced known genes
  - Confirmation and definition of hypothetical genes
  - >4500 non-coding conserved sequences in ~1700 regions that are candidates for regulatory DNA sequence element

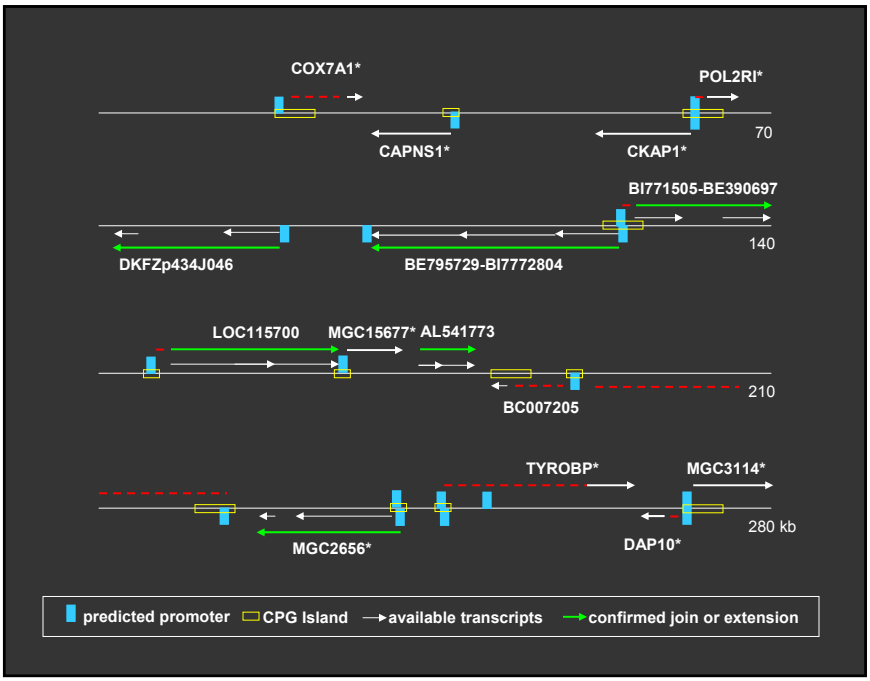
Next stage: creating a validated map of genes and associated regulatory sequences for HSA19 and mouse

- Defining the borders of known and predicted genes
  - *in silico* annotation: promoter and 5'exon prediction (with M. Zhang)
  - RTPCR, RACE to confirm gene models and define 5' and 3' ends
- Identifying and testing regulatory elements
  - Triaging promoter, enhancer candidates using high throughput reporter assays
- Linking cell-type specific expression to regulatory element structure
  - Gene expression is regulated at the level of specific cell types, but affected by tissue context
  - Can we decipher links between RE structure and specific patterns of gene expression?



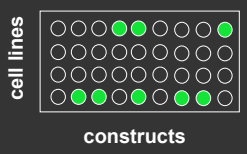

20
40
60
80
100 kb
HSA19q13.2 sequence
DNA
mus
nt
est
Refseq
protein
fly
nematode
yeast
mus
CKM
Unknown
(low % identity to cuticular collagen)
Unigene Hs.8312
(ELK kinase-like)
CKM

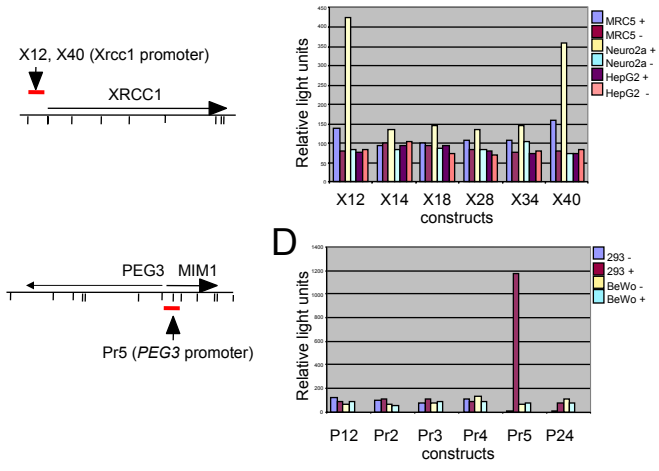

0 kb
50
100
CpG
SFBs
mus- nt
nr
EST
Refseq
promoter
Dros-x
nem-x
yeast-x
mus-x
known genes
KIAA0304 FLJ22573 MDS033
HSPOX1
gene models
200
NPHN
APLP1
DAP10 TYROBP

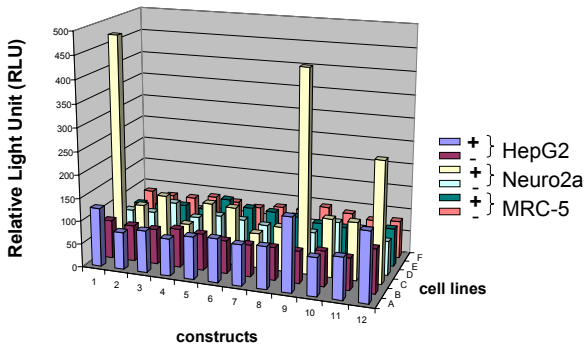


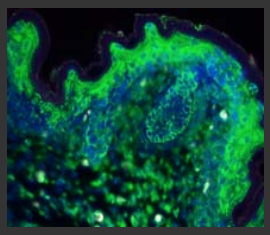
Triaging candidate sequences for regulatory function

- PCR and clone putative REs into commercially available reporter-construct vectors
- Transfect candidates into arrayed cell lines in 96 well plates, using SAGE/ microarray expression data as guide, and measure luciferase activity

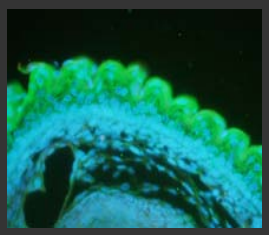



X12, X40 (Xrcc1 promoter)
XRCC1
PEG3
MIM1
Pr5 (PEG3 promoter)
Relative light units
MRC5 +
MRC5 -
Neuro2a +
Neuro2a -
HepG2 +
HepG2 -
X12 X14 X18 X28 X34 X40
constructs
D
Relative light units
293 -
293 +
BeWo -
BeWo +
P12 Pr2 Pr3 Pr4 Pr5 P24
constructs

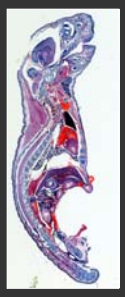

Relative Light Unit (RLU)
500
450
400
350
300
250
200
150
100
50
0
1 2 3 4 5 6 7 8 9 10 11 12
constructs
A B C D E F
cell lines
+ - HepG2
+ - Neuro2a
+ - MRC-5



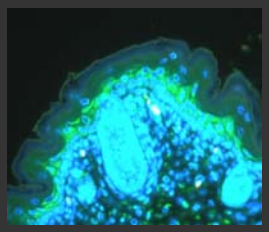
Zim1:
s. spinosum, hair bulbs,dermis



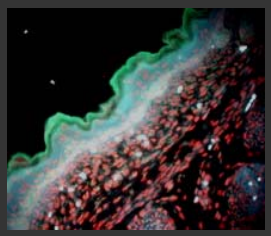
Zim2:
s. corneum, s. spinosum



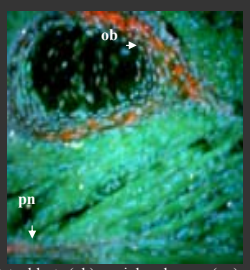
3 day-old mouse
(10 micron section)



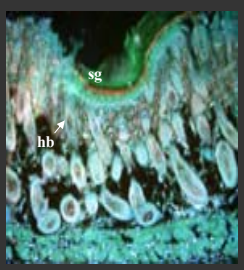
Znf264:
s. germanativum



Usp29:
s. corneum

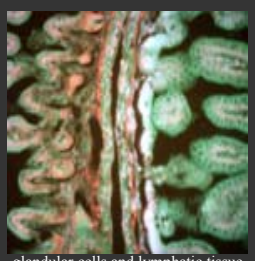


Osteoblasts (ob), peripheral nerve (pn)

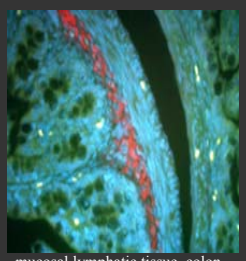


stratum granulosum (sg) and hair bulb (hb), skin

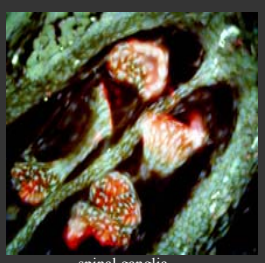
granular cells, cerebellum



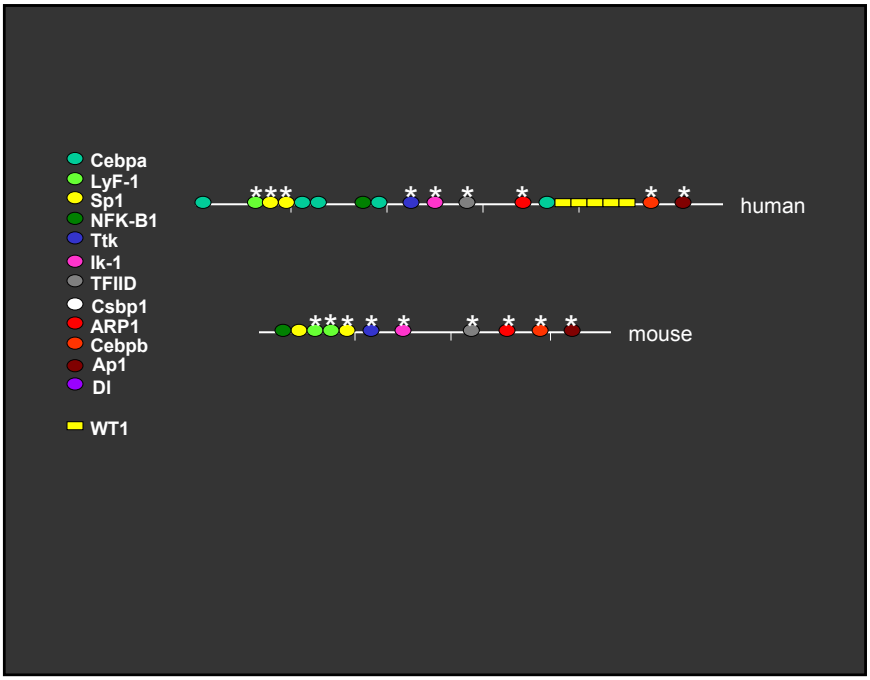
glandular cells and lymphatic tissue, stomach (left), and intestine(right)

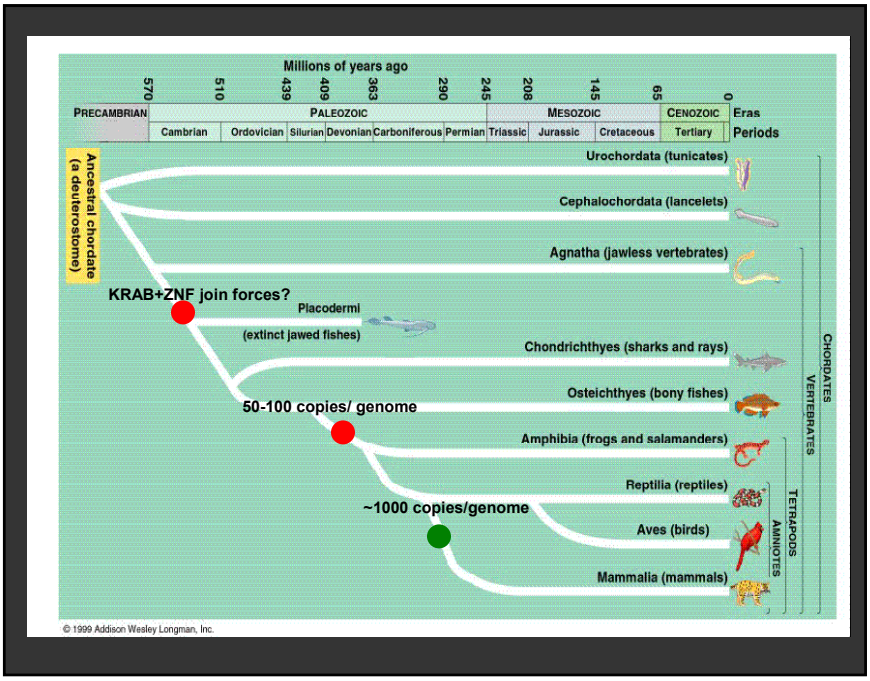


mucosal lymphatic tissue, colon



spinal ganglia


Cebpa
LyF-1
Sp1
NFK-B1
Ttk
Ik-1
TFIID
Csbp1
ARP1
Cebpb
Ap1
Dl
WT1
human
mouse


Millions of years ago
570
510
439
409
363
290
245
208
145
65
0
PRECAMBRIAN
PALEOZOIC
MESOZOIC
CENOZOIC
Eras
Cambrian
Ordovician
Silurian
Devonian
Carboniferous
Permian
Triassic
Jurassic
Cretaceous
Tertiary
Periods
Ancestral chordate (a deuterostome)
Urochordata (tunicates)
Cephalochordata (lancelets)
Agnatha (jawless vertebrates)
KRAB+ZNF join forces?
Placodermi
(extinct jawed fishes)
Chondrichthyes (sharks and rays)
Osteichthyes (bony fishes)
50-100 copies/ genome
Amphibia (frogs and salamanders)
Reptilia (reptiles)
~1000 copies/genome
Aves (birds)
Mammalia (mammals)
CHORDATES
VERTEBRATES
TETRAPODS
AMNIOTES
© 1999 Addison Wesley Longman, Inc.

many new questions:

- Are specific structures and protein binding properties associated with specific patterns of expression? if so can we discern those patterns?
- Are species-specific differences between human and mouse regulatory sequences significant, or should we pay attention only to conserved features?
- What role does regulatory element sequence variation play in human disease susceptibility, individual variation and in speciation?

Acknowledgements

- Paramvir Dehal
- Joomyeong Kim
- Laurie Gordon
- Xiaochen Lu
- Sha Hammond
- Eddie Wehri
- Angie Kolhoff

Michael Zhang, Zhenyu Xuan (CSHL)

- DNA sequencing: Elbert Branscomb,Trevor Hawkins, Paul Predki, Susan Lucas, Paul Richardson & many others (JGI)
- Sequence analysis: Art Kobayashi, Anne Olsen, Peg Folta, Astrid Terry, Carol Zhou (JGI); Ed Uberbacher, Miriam Land (ORNL)
- Mouse mappers: Anne Bergmann, Hummy Badri, Mari Christensen, Chi Ha, Mary Tran, Matt Groza, Eddie Wehri, Michelle Vargas, Mark Wagner
- Mark Shannon (ZNF family)